



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/579,235

05/30/2006

Kazuhiro Nakadai

62533.00047

8946

32294 7590 08/06/2009
SQUIRE, SANDERS & DEMPSEY L.L.P.
8000 TOWERS CRESCENT DRIVE
14TH FLOOR
VIENNA, VA 22182-6212

EXAMINER

KAZEMINEZHAD, FARZAD

ART UNIT

PAPER NUMBER

2626

MAIL DATE

DELIVERY MODE

08/06/2009

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No. 10/579,235	Applicant(s) NAKADAI ET AL.	
	Examiner FARZAD KAZEMINEZHAD	Art Unit 2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☐ Responsive to communication(s) filed on ____.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-12 is/are pending in the application.
 4a) Of the above claim(s) ____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) ____ is/are allowed.
- 6) ☒ Claim(s) 1-12 is/are rejected.
- 7) ☐ Claim(s) ____ is/are objected to.
- 8) ☐ Claim(s) ____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on ____ is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
 Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
 Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. ____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. ____. |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date ____. | 6) <input type="checkbox"/> Other: ____. |

DETAILED ACTION

Priority

1. Acknowledgment is made of applicant's claim for foreign priority under 35 U.S.C. 119(a)-(d).

Information Disclosure Statement

2. The information disclosure statement (IDS) submitted on 5/12/2006 mailing date. The submission is in compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure statement is being considered by the examiner.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1, 4, 6-7 are rejected under 35 U.S.C. 103(a) as obvious over Asano (US 2004/0054531).

Regarding claim 1, Asano does teach an automatic speech recognition system, which recognizes speeches in acoustic signals detected by a plurality of microphones as character information, the system comprising:

a sound source localization module which localizes a sound direction corresponding to a specified speaker based on the acoustic signals detected by the plurality of microphones (¶ 0129 teach the head unit 3 in Fig. 2 (performing similar

Art Unit: 2626

function as the sound source localization module) enables obtaining the direction of sound utilizing a plurality of microphones at a target (e.g. a robot) which receives the speech signals by computing power and phase differences of speech signals due to sources attributed to users (speakers); Abstract teaches all incoming speech undergo speech recognition utilizing plural sets of acoustic models);

a feature extractor which extracts features of speech signals contained in one or more pieces of information detected by the plurality of microphones (the feature extractor unit 101 in Fig. 9 receives speech data from microphones (unit 21 Fig. 9) via the analog to digital converter. ¶ 0104 lines 1-4 teach extracting feature vectors of the incoming speech data) ;

an acoustic model memory which stores distance-dependent acoustic models that are adjusted to a plurality of distances at intervals (¶ 0131 lines 4-8 referring to Fig. 9 disclose storing acoustic models corresponding to selected distances in the database units (104)_1 (104)_N (located in the memory module 42 in Fig. 3); these data are used following determination of distance of a sound source attributed to a user; ¶ 0106 teaches all feature vectors corresponding to acoustic analysis (¶ 0104 lines 1-2) are stored in speech periods (intervals) corresponding to an utterance);

an acoustic model composition module which composes an acoustic model adjusted to the sound direction, which is localized by the sound source localization module, based on the distance-dependent acoustic models in the acoustic model memory, the acoustic model composition module storing the acoustic model in the acoustic model memory (¶ 0114 teaches "N" acoustic models corresponding to "N"

Art Unit: 2626

sound sources located at “N” distances are produced (composed) where each acoustic model corresponding to a certain distance (localization) is stored in a certain database (e.g. one of the units (104)_1 to (104)_N in Fig. 9));

and a speech recognition module which recognizes the features extracted by the feature extractor as character information using the acoustic model composed by the acoustic model composition module (§ 0132 (lines 2-1 above § 0133) teaches module units 41B and 41A in Fig. 3 as disclosed in step S4 in Fig. 10 enable performing speech recognition utilizing feature vectors extracted from the speech data (§ 0132 lines 1-3); § 0108 lines 1-4 teach acoustic model databases used for the speech recognition have stored acoustic characteristics of phonetic-linguistic-units such as phonemes or syllables (e.g. character units comprising words) in them) .

Asano does not specifically disclose storing direction-dependent acoustic models in its acoustic model memory for time intervals corresponding to speech signals. It would have therefore been obvious to one with ordinary skill in the art at the time the invention was made to not only store distance-dependent acoustic models in the memory as is done here in modules 104 in Fig. 9, but also direction-dependent acoustic models obtained by using the stored distance-dependent acoustic models in obtaining the direction of sound using the methods of § 0129 so that the robot will have bias for direction as well as distance and will point at the direction of the source of sound to reduce detecting wrong signals and enhance its speech recognition performance.

Art Unit: 2626

Regarding claim 4, Asano does teach a system according to claim 1, wherein the sound source localization module employs scattering theory that generates a model for an acoustic signal, which scatters on a surface of a member (¶ 0142 lines 4-9 and ¶ 0145 teach using reflected (scattered) ultrasonic wave pulses from an obstacle, the robot can determine the distance of a user producing speech from the robot which according to ¶ 0131 is used in generating distance dependent acoustic models; all these processes are made possible by the sensor unit 111 (Fig. 11) attached to the robot);

specifying the sound direction for the speaker with the intensity difference and the phase difference detected from the plurality of microphones (¶ 0129 lines 1-5);

Asano does not specifically teach using scattered (reflected) waves from the surface to which the microphones are attached to determine the sound direction attributed to a user (speaker). But it would have been obvious to one with ordinary skill in the art at the time the invention was made to utilize the sensor unit 111 to create a second model in which one determines the phase and power difference of the sound waves reflected (scattered) from the surface of the robot in determining the direction of the speech signals and thereby create a second model of determination of direction of a sound wave incident on the robot by analyzing reflected rather than incident waves on the robot and thereby help in generating more accurate sound direction by benchmarking the results of the two models against one another.

Art Unit: 2626

Regarding claim 6, Asano does not specifically disclose a system according to claim 1, wherein the acoustic model composition module is configured to compose an acoustic model for the sound direction by applying weighted linear summation to the direction- dependent acoustic models in the acoustic model memory, and weights introduced into the linear summation are determined by training.

Asano however does teach using N stored distance dependent acoustic models (104 units in Fig. 9) to select one acoustic model among them which has the closest match for a sound source which is located within the range of the distances corresponding to the said acoustic models by using its matching unit 103 which does the matching by calculating acoustic scores (§ 0122). § 0121 lines 1-4 teaches distance calculator (unit 47 in Fig. 3) enabling calculation of distance of the robot from a user uttering speech (speaker).

It would have therefore been obvious to one with ordinary skill in the art at the time the invention was made to compose an acoustic model corresponding to a sound source using a linear interpolation of acoustic models which correspond to the distances closest to the said sound source and train the acoustic model by changing the coefficients of the linear interpolation to achieve the best score, and thereby compose a more realistic acoustic model for the said sound source resulting in better performance by the robot.

For extension of this discussion from distance dependent to direction dependent acoustic models please see the obviousness of the claim 1.

Regarding claim 7, Asano does teach a system according to claim 1, further comprising a speaker identification module, wherein the acoustic model memory possesses the direction-dependent acoustic models for respective speakers (identical to claim 1, 3rd limitation and rejected under similar rationale),

and wherein the acoustic model composition module is configured to execute a process comprising: referring to direction-dependent acoustic models of a speaker who is identified by the speaker identifying module and to a sound direction localized by the sound source localization module (¶ 0116 teaches the acoustic model database to have the acoustic model of speakers (specific speakers) at specified locations; ¶ 0159 teaches those acoustic models are produced by from the speech data acquired by microphone placed close to the mouth of the speakers for better voice recognition leading to better speaker identification; ¶0130 lines 5-1 above ¶0131 teach image of the face of a user (e.g. a potential speaker) taken by the robot's CCD cameras (modules 22L and 22R in Fig. 2) are used as a reference pattern (stored) for image recognition (user identification); finally ¶ 0145 lines 1-5 teach the robot uses both detection of a user uttering (identifying a user by his voice) as well as image recognition (identifying a user by his image) in orienting his head unit in the direction of the user; i.e., speaker identification is done by both image pattern recognition as well as by voice recognition using the stored acoustic models of the user (speaker));

composing an acoustic model for the sound direction based on the direction-dependent acoustic models in the acoustic model memory; and storing the acoustic

Art Unit: 2626

model in the acoustic model memory (corresponds to the third limitation of the first claim and rejected under similar rationale).

For obviousness analysis please see claim 1.

5. Claim 2-3, 8-12 are rejected under 35 U.S.C. 103(a) as being unpatentable over Asano, and further in view of Ito et al. (US Patent 7,076,433).

Regarding claim 2, Asano does teach an automatic speech recognition system, which recognizes speeches of a specified speaker in acoustic signals detected by a plurality of microphones as character information, the system comprising:

a sound source localization module which localizes a sound direction corresponding to the specified speaker based on the acoustic signals detected by the plurality of microphones (identical to the first limitation of claim 1 and rejected under similar rationale);

an acoustic model memory which stores distance-dependent acoustic models that are adjusted to a plurality of directions at intervals (identical to the 3rd limitation of claim 1 and rejected under similar rationale) ;

an acoustic model composition module which composes an acoustic model adjusted to the sound direction, which is localized by the sound source localization module, based on the distance-dependent acoustic models in the acoustic model memory, the acoustic model composition module storing the acoustic model in the acoustic model memory (identical to the 4th limitation of claim 1 and rejected under similar rationale) ;

and a speech recognition module which recognizes the features extracted by the feature extractor as character information using the acoustic model composed by the acoustic model composition module (identical to the 5th limitation of claim 1 and rejected under similar rationale)

Asano does not specifically disclose a sound source separation module which separates speech signals of the specified speaker from the acoustic signals based on the sound direction localized by the sound source localization module

a feature extractor which extracts features of the speech signals separated by the sound source separation module;

and storing direction-dependent acoustic models in its acoustic model memory for time intervals corresponding to speech signals.

Ito et al. does teach a sound source separation apparatus which in one embodiment separates sound (e.g. attributed to a speaker) from a mixed input signal (acoustic signal) by utilizing sound source direction as an acoustic feature (Abstract lines 1-2 first ¶ and line 4 from the bottom; Col. 18 lines 61 to 66 referring to Fig. 16 unit 911). Furthermore it teaches its acoustic feature extractor to incorporate a sound source direction prediction layer which aids in determination of the peaks corresponding to acoustic features of the sound source incident from a certain direction (Col. 19 lines 33-40 and module 921 in Fig. 16).

It would have therefore been obvious to one with ordinary skill in the art at the time the invention was made that utilizing the modules 915 and 921 in Fig. 16 of Ito et al. into the feature extractor unit 101 of Fig. 9 of Asano would enable the latter to

Art Unit: 2626

separate a sound (e.g a speech signal) from a mixed input when the sound is incident from a certain direction by utilizing its direction dependent features enabling the robot to obtain bias for direction for a certain speaker and will point at the direction of the source of sound to reduce detecting wrong signals and enhance its speech recognition performance. Storage of direction dependent features also aids in avoiding calculations of the said features and helps in raising efficiency.

Regarding claim 3, Asano does suggest a system according to claim 1, wherein the sound source localization module is configured to execute a process comprising:

acquiring an intensity difference and a phase difference for the harmonic relationships extracted through the plurality of microphones (¶ 0129 lines 1-5 teach acquiring power (intensity) and phase difference of speech signals (which maintain harmonic spectra) picked up by a device (e.g. robot) microphones);

acquiring belief factors for a sound direction based on the intensity difference and the phase difference, respectively; and determining a most probable sound direction (utilizing power and phase difference in determining the direction of sound as disclosed in ¶ 0129 lines 1-5 does inherently involve these or equivalent steps to achieve the same net outcome (sound direction determination)).

Asano does not specifically disclose performing a frequency analysis for the acoustic signals detected by the microphones to extract harmonic relationships.

Ito et al. does disclose performing a frequency analysis for the acoustic signals detected by the microphones to extract harmonic relationships (Col. 12 lines 45-50

Art Unit: 2626

teach harmonic calculation layer (named as the intermediate feature extraction layer unit 107 in Fig. 9) determines harmonic features of features (attributed to acoustic analysis of speech signals) at each time based on their frequency variation rates (i.e. by doing frequency analysis));

It would have therefore been obvious to one with ordinary skill in the art at the time the invention was made that utilizing the harmonic calculation layer (unit 107 in Fig. 9) into the feature extractor unit 101 of Fig. 9 of Asano would enable Asano to extract harmonic structures attributed to speech signals incident on the microphones of the robot prior to obtaining their phase and power differences for the sake of determining the direction of the speech and thereby eliminate redundant parts of the spectra in determining direction of speech which primarily included harmonics and thereby enhance efficiency and accuracy.

Regarding claim 8, the preamble, the 1st, 4th, 5th, 6th and 7th limitations correspond to the preamble, 1st, 2nd, 3rd, 4th and 5th limitations respectively of the claim 1 and are therefore rejected under similar rationale over Asano.

Limitation 3, corresponds to the limitation 2 of the claim 2 and is therefore rejected under similar rationale by Ito et al.

Regarding the 2nd limitation, storing direction dependent acoustic models at time intervals (the 3rd limitation of claim 1) will amount to storing the sound direction at time intervals corresponding to the speech attributed to a sound source and the storage medium used (i.e. the memory unit 42 in Fig. 3) will enable the unit the functionality of

Art Unit: 2626

the stream tracking module claimed in this limitation which will enable estimating the current position of a sound source by simple interpolation.

For obviousness please see claims 1 and 2.

Regarding claims 9, 10, 11, 12 they correspond to claims 3, 4, 6, 7 respectively with identical limitations and are therefore rejected under similar rationales.

6. Claim 5 rejected under 35 U.S.C. 103(a) as being unpatentable over Asano in view of Ito et al., and further in view of Okuno et al. (US Patent 7,035,418).

Regarding claim 5, Asano in view of Ito et al. does not specifically disclose a system according to claim 2, wherein the sound source separation module employs an active direction-pass filter so as to separate speeches, the filter being configured to execute a process comprising:

separating speeches by a narrower directional band when a sound direction, which is localized by the sound source localization module, lies close to a front, which is defined by an arrangement of the plurality of microphones;

and separating speeches by a wider directional band when the sound direction lies apart from the front.

Okuno et al. does teach a directional filter for sound based on the position of the source, enabling localization and extracting sound (speech) information of the said sound source which will thereby enable it to separate that source from other sound

Art Unit: 2626

(speech) sources (Col. 4 lines 25-34); Col. 8 lines 1-7 referring to the flow chart on Fig. 8 identifies steps ST5 and ST6 as the steps associated with the directional filter's operation; Col. 8 lines 15-20 note the results of directional filter on detecting directions of sound from three sources (A, B and C on Fig. 7) and notes that the angular range (directional band) about which these directions are determined to have been reduced due to the application of the filter; Col. 11 lines 48-52 teach the directional filter functions according to the direction and position of the sound source which is determined first by computing the difference between phase and intensity of sound received at two receivers from the same source which is directly related to their path differences; Col. 6 lines 32-35 referring to Fig. 4 explicitly show a relationship ($d = D \sin(\theta)$) between the two "sound" signal path differences (attributed to the source position which is called "d"), the distance between the receiving microphones (called "D") and the angle governing the direction of the sound source with respect to an axis at right angles to the line connecting the two microphones (called "theta") which shows the angle and thereby the angular range (directional band) to increase with that path difference.

It would have therefore been obvious to one with ordinary skill in the art at the time the invention was made that utilizing these methods (i.e., steps ST5 and ST6 in Fig. 8 of Okuno et al.) for directional filter into the flow chart of Fig. 10 of Asano (to after step S2) would enable the latter to apply the distance dependent directional filter of Okuno et al. which would result in less accuracy and larger angular range (wider directional band) for further sound sources since the incoming signals possess longer

Art Unit: 2626

path differences as larger path difference (i.e. larger "d" in the relationship above) lead to larger angular range (i.e. larger deviations (directional band) in "theta"). Application of directional filter will reduce the error in determining the direction of the sound source in general.

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Maekawa et al. (US Patent 6,471,420), Almstrand et al. (US 2003/0229495).

Any inquiry concerning this communication or earlier communications from the examiner should be directed to FARZAD KAZEMINEZHAD whose telephone number is (571)270-5860. The examiner can normally be reached on M-F 8:30AM-5:00 PM EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis I. Smits can be reached on (571)272-7628. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2626

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FK/

/Talivaldis Ivars Smits/
Primary Examiner, Art Unit 2626

8/4/2009